Volume 9 Number 13 1981 Nucleic Acids Research

Shotgun DNA sequencing using cloned DNase I-generated fragments

Stephen Anderson

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received 7 May 1981

ABSTRACT

A method for DNA sequencing has been developed that utilises libraries of cloned randomly-fragmented DNA. The DNA to be sequenced is first subjected to limited attack by a non-specific endonuclease (DNase I in the presence of Mn^{++}), fractionated by size and cloned in a single-stranded phage vector. Clones are then picked at random and used to provide a template for sequencing by the dideoxynucleotide chain termination method. This technique was used to sequence completely a 4257 bp EcoRI fragment of bovine mitochondrial DNA. The cloned fragments were evenly distributed with respect to the EcoRI fragment, and completion of the entire sequence required the construction of only a single library. In general, once a clone library has been prepared, the speed of this approach (>1000 nucleotides of randomly selected sequence per day) is limited mainly by the rate at which the data can be processed. Because the clones are selected randomly, however, the average amount of new sequence information per clone is substantially diminished as the sequence nears completion.

INTRODUCTION

A large DNA fragment can be completely sequenced by cloning smaller sub-fragments in a single-stranded phage vector to produce a representative library, then sequencing randomly-chosen clones using the dideoxynucleotide chain termination method and a flanking universal primer [1-6]. This method, in which the final sequence is built up as a composite of overlapping sub-fragment sequences, has been aptly termed "shotgun" DNA sequencing. Recent advances in computer software that allow the efficient handling and comparison of large amounts of sequence data [7-9] have made this strategy especially attractive for the complete sequence determination of long stretches of DNA.

An optimal library for DNA sequencing should contain clones of fragments having overlapping sequences, and clones representative of each part of the DNA region of interest should be present in roughly equal proportions. Furthermore, although DNA fragments larger than about 1000 bp cannot always be stably inserted into the single-stranded phage vectors such as M13mp2 [6,

10-12], insertions greater than 350 bp in length are desirable in order to take full advantage of the resolving power of DNA sequencing gels. Thus, the process used to produce the DNA fragments to be inserted should be capable of generating DNA fragments falling predominantly within these size limits.

Restriction endonucleases have been used previously to generate fragments for shotgun DNA sequencing [4,6], but their general utility in this regard is limited by several factors. Because individual restriction fragments produced by a given enzyme will not yield sequences that meaningfully overlap, and because restriction enzyme sites are usually distributed non-uniformly in complex DNA molecules, it is necessary to prepare a number of clone libraries, each based upon a digestion with a different restriction enzyme, in order to establish a complete sequence. Most fragments produced by restriction enzymes with four-base recognition sites, moreover, are shorter than the length of accurate sequence that can be read from a gel, so sequencing such fragments represents an inefficient use of resources.

For the above reasons a simple method was developed for shotgun DNA sequencing using cloned DNA fragments generated by the action of pancreatic DNase (DNase I) in the presence of Mn^{++}. To test the effectiveness of the method it was used to determine the complete sequence of a 4257 bp EcoRI fragment of bovine mitochondrial DNA containing several known genes for mitochondrial tRNAs and cytochrome c oxidase subunits (S. Anderson, M.H.L. de Bruijn, A.R. Coulson, I.C. Eperon, F. Sanger, R. Staden and I.G. Young, manuscript in preparation).

MATERIALS AND METHODS

DNase digestion and size fractionation of fragments

DNase I was stored in aliquots at -20°C as a 1 mg/ml solution in 0.01 N HCl [13] and diluted one hour before use in digestion buffer (see below). DNase I stored in this manner was found to lose activity with a half-life of approximately three months, and lot-specific variations in activity have been reported [13], so in general it was necessary to do a series of test digestions over a range of enzyme concentrations before each experiment. Extents of digestion were monitored by agarose gel electrophoresis of the resultant fragments and a DNase I concentration was chosen that would produce a distribution of fragments with a median size of 300-500 bp.

A purified ^{32}P-labelled 4257 bp EcoRI restriction fragment of bovine mtDNA, derived from plasmid pBMT57 [14], was first circularised with T4 DNA ligase. After phenol extraction and concentration by ethanol precipitation

6 µg of this DNA was incubated with 1.5 ng of DNase I for 10 min at 15°C in a freshly prepared buffer composed of 50 mM Tris-HCl (pH 7.5), 1.0 mM $MnCl_2$, 100 µg/ml bovine serum albumin, in a final volume of 90 µl; the digestion was stopped by the addition of 5 µl of 100 mM EDTA.

The DNA was extracted with phenol:chloroform and with ether, then sedimented on a 5-20% neutral sucrose gradient in a Beckman SW60 rotor at 50,000 rpm for 11 hr at 4°C. 100 µl fractions were collected and the average size of the fragments in each fraction determined by agarose gel electrophoresis of aliquots. Fractions representing fragments >300 bp long (approximately 50% of total DNA) were pooled and the DNA precipitated with ethanol.

Repair and cloning of DNA fragments

Nicks and potential gaps in the DNA were repaired as described previously [15]. In brief, 2 µg of DNA fragments were treated, in a final volume of 50 µl, with 10 units of T4 DNA polymerase and 5 units of *E. coli* DNA ligase for 60 min at 15°C in a buffer composed of 40 mM Tris-HCl (pH 7.9), 10 mM $(NH_4)_2SO_4$, 10 mM β-mercaptoethanol, 5 mM $MgCl_2$, 0.5 mM EDTA, 100 µM each dATP, dCTP, dGTP and dTTP, and 50 µM NAD^+. After stopping the reaction with 5 µl of 100 mM EDTA the DNA was extracted with phenol:chloroform and with ether, then concentrated by ethanol precipitation.

"Linker" oligodeoxynucleotides containing a *BamHI* restriction site (C-C-G-G-A-T-C-C-G-G) were attached to the repaired DNA as described by Maniatis *et al.* [16], except that one unit of *E. coli* DNA polymerase I large fragment and 200 µM each dATP, dCTP, dGTP and dTTP were included in the ligation reaction in order to increase the proportion of flush ends on the DNA fragments. Linker oligodeoxynucleotides were present during the ligation at a 50-fold molar excess over DNA fragments. Ligations were terminated by adding EDTA to a final concentration of 20 mM and the DNA was extracted with phenol:chloroform and with ether, then concentrated by ethanol precipitation. The ligated DNA-linker mixture was incubated with 30 units of *BamHI* per µg of linker for 120 min at 37°C, then passed over a 1.5 ml column of Ultrogel AcA34 equilibrated with 10 mM Tris-HCl (pH 7.8), 1 mM EDTA, 400 mM NaCl. The digested excess linkers were found to bind very tightly to this column, whereas all DNA fragments larger than 50 bp eluted in the void volume. The purified DNA fragments were cloned using an amber derivative of the vector M13mWJ22 [17] in thebacterial host JM101 [Δ(*lacpro*) *supE thi*, F' *traD36 proAB lacI*^{q}, ZΔM15] [18]. These experiments were approved for EK1-C1* containment by the Genetic Manipulation Advisory Group and were carried out

under such conditions in accordance with the NIH Guidelines.

Preparation of templates and sequencing

Plaques representing recombinant clones were identified by their Lac^- phenotype on Xgal indicator plates [10,11], picked at random, and grown for 10 hr at 37°C in 1 ml of 2 x TYE broth containing 0.01 ml of a fresh JM101 overnight culture. After a brief centrifugation to remove the cells, the phage were precipitated from the medium using polyethylene glycol 6000 [19]. The precipitated phage were dissolved at 55°C in 50 µl of 10 mM Tris-HCl (pH 7.8), 1 mM EDTA, 0.2% Sarkosyl, 50 µg/ml proteinase K, then extracted with phenol:chloroform and with ether. The DNA was precipitated with ethanol and redissolved in 50 µl of 10 mM Tris-HCl (pH 7.8), 1 mM EDTA.

The DNA was sequenced by the dideoxynucleotide chain termination method using either a cloned primer from plasmid SP14 [3] or a 17-nucleotide long synthetic primer [20]. Annealing and sequencing reactions were carried out as previously described [3] on four clones at a time, except that [α-^{32}P]dATP was used as the source of label and 0.05 units of E. coli DNA polymerase I large fragment were included in the "chase" solution. The sequencing reactions were terminated by the addition of 6 µl of 98% deionised formamide, 10 mM EDTA, 0.2% bromphenol blue and 0.2% xylene cyanol.

After heating the terminated reaction mixtures for 6 min at 100°C, 2 µl aliquots were loaded onto 6% polyacrylamide-7 M urea "thin" gels [21] and electrophoresed for 1.5 or 4 hr at 25-30 mA (1.3-1.7 kV). The 1.5 hr gel was made up and run in 90 mM Tris, 90 mM boric acid, 2.5 mM EDTA (pH 8.3), while the 4 hr gel was made up and run in 135 mM Tris, 45 mM boric acid, 2.5 mM EDTA (pH 8.9); the latter buffer was used for the longer runs because it was found to yield sharp bands more reproducibly. After electrophoresis gels were fixed in 10% acetic acid and autoradiographed at room temperature.

Materials

DNase I (grade DN-EP) was from Sigma. E. coli DNA polymerase I large fragment, T4 DNA polymerase, NAD^+, and deoxynucleoside triphosphates were from P-L Biochemicals; BamHI linkers and dideoxynucleoside triphosphates were from Collaborative Research. E. coli DNA ligase and BamHI were from New England Biolabs, Ultrogel AcA34 from L.K.B. and [α-^{32}P]dATP and [γ-^{32}P]rATP from the Radiochemical Centre, Amersham. T4 DNA ligase was a gift from Dr. A.R. MacLeod, and an amber derivative of the vector M13mWJ22 was a gift from Dr. G.P. Winter. E. coli strain JM101 was provided by Dr. J. Messing.

RESULTS

Properties of the DNase digestion

DNase I produces, in the presence of Mn^{++}, double-stranded breaks in DNA [22] and has been reported to have only limited sequence specificity in its cutting [23-25]. In order to assess the feasibility of using this enzyme to produce randomly fragmented DNA of arbitrary size for cloning purposes, the properties of DNase I digestion in the presence of Mn^{++} were explored further. The fragmention pattern was tested by digesting samples of lambda DNA with various concentrations of DNase I and separating the products by agarose gel electrophoresis. The average sizes of the resultant fragments were found to be inversely proportional to the amount of enzyme in the digest and the distributions were smooth and unimodal (Fig. 1). This indicated that there were no obvious preferentially digested or resistant sequences present in the digest.

DNA fragments produced by DNase I in the presence of Mn^{++} have been reported to have staggered ends [27,28]. Cloning of the fragments thus required that the ends be treated to render them flush. Since the choice of the appropriate enzymatic treatment depended on whether or not internal nicks were present in the fragments, experiments were performed to determine the ratio of single strand nicks to duplex cuts produced by DNase I in the presence of Mn^{++}. The sedimentation behaviour of the digested DNA on neutral and on alkaline sucrose gradients in the presence of restriction fragments as size standards (not shown) indicated that an average of 2.7 internal phosphodiester bond interruptions were present per fragment. A similar value was obtained when the total number of 5'-termini present in the digested DNA was measured, using [γ-^{32}P]rATP in a polynucleotide kinase-mediated exchange reaction [29], before and after a nick repair treatment with *E. coli* DNA ligase (not shown). At dilute enzyme concentrations this ratio was independent of the extent of digestion, and thus under the conditions used DNase I appeared to make in the DNA approximately 2.7 nicks for every duplex cut.

Cloning and sequencing of DNase I-generated fragments

The usefulness of a DNase I/Mn^{++} digestion in the creation of uniformly representative clone libraries for shotgun DNA sequencing was tested with a 4257 bp *Eco*RI fragment from the bovine mitochondrial genome [14]. This fragment was circularised by ligation of its cohesive termini then digested as described in Materials and Methods. The resulting DNA fragments were subjected to a size fractionation by velocity sedimentation on a neutral sucrose gradient, and those fractions containing DNA fragments in the range

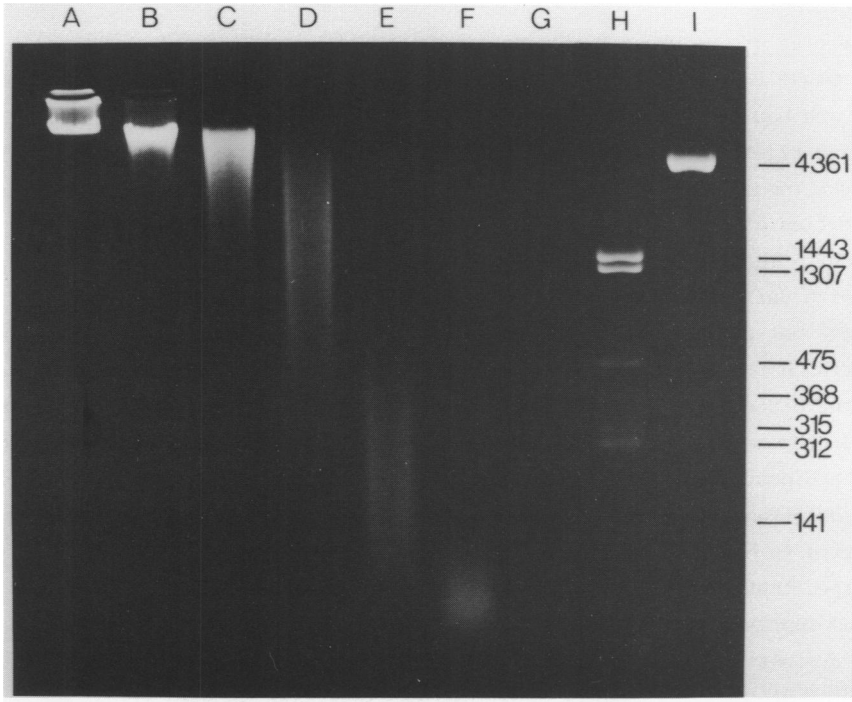

Figure 1. Size distribution of DNA after DNase I digestion. 3 µg samples of λ DNA were digested with various amounts of DNase I in the presence of Mn^{++} as described in Materials and Methods. Aliquots of these samples were electrophoresed on a 1.8% agarose gel and stained with ethidium bromide. Included on the same gel were TaqI and EcoRI restriction fragments of pBR322 [26] as size markers. The amount of DNase I used for the digestion associated with each lane was as follows: A, none; B, 0.022 ng; C, 0.075 ng; D, 0.22 ng; E, 0.75 ng; F, 2.2 ng; G, 7.5 ng. Lanes H and I represent TaqI- and EcoRI-digested pBR322 DNA, respectively; the sizes (in bp) of the restriction fragments are given in the margin.

300-1000 bp were pooled. To repair the nicks or any small gaps the DNA fragments were treated with E. coli DNA ligase plus T4 DNA polymerase [15,30], and to ensure that a high proportion of the fragments would have flush ends the DNA was subsequently treated with E. coli DNA polymerase I large fragment [3,4].

In order to clone the DNase I-generated fragments synthetic oligodeoxynucleotide "linkers" were attached to the flush ends using T4 DNA ligase. BamHI linkers were chosen because the bovine mtDNA EcoRI fragment was known to contain no BamHI sites [14]. After cleavage with BamHI to produce cohesive

termini the fragments were cloned into the BamHI site of M13mWJ22(amber) [17], a derivative of M13mp2 [10,11]. The overall yield of recombinant phage, detected as colourless (Lac^-) plaques on Xgal indicator plates, was greater than 2×10^3 pfu per μg of the EcoRI fragment. Clones were picked at random and their respective inserts sequenced as described in Materials and Methods.

Sequence assembly

Cloned inserts were sequenced at a rate of four per day and the sequences compiled by computer using the programs BATIN, CONSEN, DBCOMP and DBUTIL [9]. Initially 48 Lac^- clones were picked and sequenced; 46 of these proved to have inserts of bovine mitochondrial DNA and were sufficient to cover completely the 4257 bp EcoRI fragment. However, in order to confirm several regions of the sequence 36 more cloned templates were prepared and subjected to sequencing reactions in which ddTTP was the only chain terminating nucleotide analogue present. These partial sequences ("T tracks"), showing only the order of T and non-T residues, were sufficient to locate the positions of the clones in the nearly complete master sequence. Eight of these clones, representing regions where confirming sequences was desired, were used to complete and verify the sequence of the 4257 bp EcoRI fragment (Fig. 2). This sequence exhibited greater than 70% homology with the sequence of the corresponding region of the human mtDNA molecule [31] and will be presented elsewhere as part of the complete sequence of the bovine mitochondrial genome (S. Anderson, M.H.L. de Bruijn, A.R. Coulson, I.C. Eperon, F. Sanger, R. Staden and I.G. Young, manuscript in preparation).

Clone distribution

Figure 2 shows that the sequenced cloned fragments were distributed fairly uniformly with respect to the 4257 bp EcoRI fragment, indicating that no portion of the sequence was highly over- or under-represented in the clone population. Information for 98% of the nucleotide positions in the sequence was obtained from at least two independent clones, and 75% of the sequence was represented by clones from both complementary strands.

The sizes of the bovine mtDNA fragments inserted into the vector ranged from 88 to greater than 370 bp (the limit of readable sequence on the 4 hr gel), even though the neutral sucrose gradient purification was intended to exclude fragments less than 300 nucleotides in length. This discrepancy may be due to the relatively low resolution of size fractionations on neutral sucrose gradients or to juxtaposed nicks in opposite strands allowing adventitious breakdown of the DNA fragments subsequent to the sedimentation step. The representation of any contaminating small fragments would be

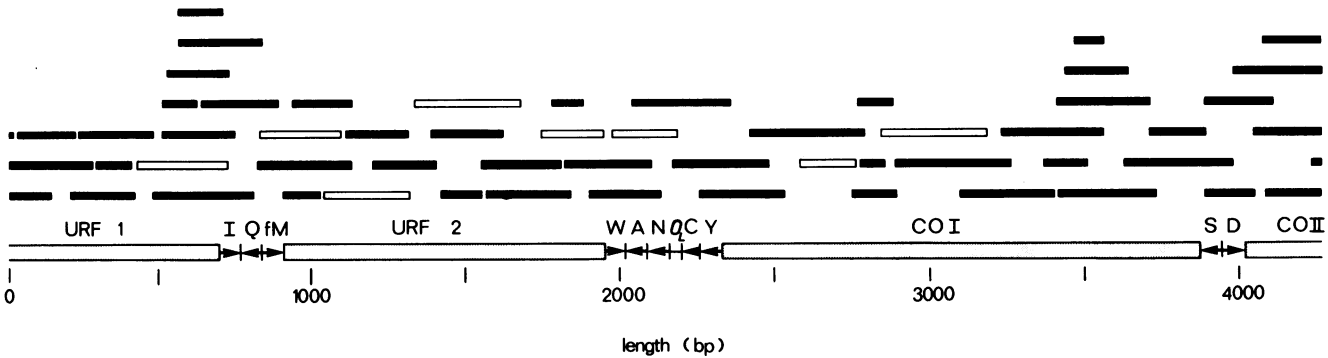

Figure 2. Sequence distribution of cloned DNase I fragments. The sequences of the cloned inserts that were overlapped to produce the complete sequence of the 4257 bp EcoRI fragment are depicted as bars and are positioned to show the contribution of each clone to the final sequence. Solid bars represent sequences from the initial random selection of 48 clones; open bars represent the eight confirming sequences that were done after T-track screening of 36 additional clones (see text). Shown schematically are several bovine mitochondrial genes identified from the DNA sequence of the EcoRI fragment [31]. These include the coding regions for cytochrome c oxidase subunits I and II (COI and COII) and two other large unidentified reading frames (URF1 and URF2) that presumably also code for proteins. Genes for mitochondrial tRNAs are labelled according to the one-letter amino acid code and are depicted as → or ← depending on whether the tRNAs have the sense of the L- or the H-strand, respectively. O_L is the presumptive origin of L-strand synthesis during mtDNA replication.

accentuated in the library because of the substantially greater efficiency with which they are cloned (S. Anderson, unpublished results).

DNase I cutting specificity

Neither the T4 DNA polymerase nor the E. coli DNA polymerase I large fragment used to treat the DNA will attack 5'-termini at the ends of the DNaseI-generated duplex fragments. The nucleotides comprising these termini were thus preserved in the clones, and from the sequences at the ends of the cloned inserts it was possible to locate precisely the sites of DNase I cutting. The "context" of each cut site, arbitrarily defined as the 20-nucleotide-long stretch of DNA centred on the site, was examined in order to detect any sequence specificity in the cutting. The combined data from 92 sequenced cut sites indicated that there was very little base composition bias in the region of cutting; cut site contexts defined as above had an average base composition of 39.3% G + C, whereas the composition of the entire fragment was 39.2% G + C. There did appear to be some degree of sequence recognition by the enzyme, however, in that 33% of the cut sites occurred one nucleotide to the 5' side of the dinucleotide sequence G-T. This is more than three times the frequency that would be expected if the cuts were located randomly with respect to G-T dinucleotides.

Sequence accumulation

The total accumulated unique sequence was plotted as a function of the number of clones sequenced in order to show the rate at which sequence information can be acquired with this method (Fig. 3). The plotted points closely match the theoretical accumulation curve expected for a process in which sequence information is acquired in a totally random fashion. However, as the project was nearing completion the points converged on the complete sequence more rapidly than the theoretical curve would predict. The significance of this, given only one example, is unclear.

DISCUSSION

This paper describes the production of uniformly representative clone libraries based on a single-stranded phage vector and demonstrates the use of such a library for shotgun DNA sequencing by the dideoxynucleotide chain termination method. The cloned fragments that make up the library were produced by the virtually sequence-independent digestion of DNA by DNase I in the presence of Mn^{++}. Similar procedures have been used previously for the in vitro production of randomly placed deletions [32,33] or insertions [28] in DNA. The cloning method outlined in this paper involved a "pre-repair" of

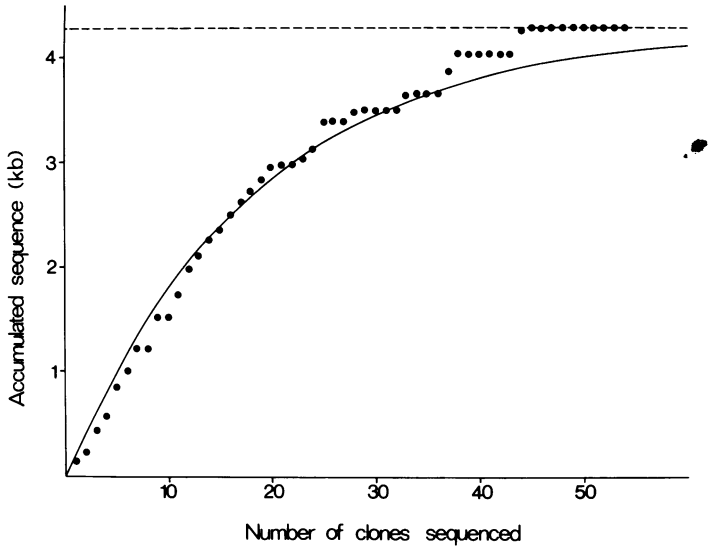

Figure 3. Rate of sequence acquisition. The points represent a plot of the total amount of determined sequence (in kilobases) from the EcoRI fragment versus the total number of clones sequenced at each stage in the project. The broken line extending from the ordinate demarcates the 4257 nucleotide length of the completely determined sequence; the last eight points lying on this line represent those clones that were sequenced for confirmation purposes. The solid line is a continuous plot of the discrete function $S(n) = L[1-(1-\frac{\bar{l}}{L})^n]$, where S(n) = total accumulated sequence, n = total number of clones sequenced, L = 4257 and $\bar{l}$ = 227, the average size of the insert sequences. This function represents a summation of the expected amounts of new sequence data obtainable from successive clones if it is assumed they have been chosen at random from a uniformly representative library. Such a library is defined as being one for which the probability of picking a clone representing any portion of the target sequence is uniform.

nicks in the DNA fragments followed by treatment with E. coli DNA polymerase I large fragment to create flush ends. More recent experiments (S. Anderson, unpublished) have shown that a single treatment with T4 DNA polymerase at 11°C [34] is sufficient for this purpose. The use of linkers may also be avoided by direct ligation into a flush-ended insertion site such as the HincII site in the vector M13mp7 [6].

Shotgun DNA sequencing using cloned randomly-fragmented DNA is a very rapid and accurate procedure provided certain precautions are observed. For example, the joining of two or more randomly-cut DNA fragments via blunt-end ligation may occur during the cloning steps; unlike the case in which two

restriction fragments are joined together, there would be no clues from the DNA sequence as to where such joints might have been formed. This problem can be obviated by cloning DNA fragments of sufficient size (e.g. >500 bp) such that the extent of readable sequence obtained using a flanking primer is always less than the fragment length. A second safeguard is to attach oligodeoxynucleotide linkers to the DNA using ligation conditions under which linkers are in large excess, thus making it likely that any fragment-fragment joints will be marked by the presence of a linker sequence. The best precaution, however, is to derive each portion of the sequence from at least two independent clones.

Once a library has been constructed and template DNA prepared from the clones it is possible to generate sequence at a rate of more than 1000 nucleotides per day. If only the biochemical manipulations were considered this rate could be readily doubled or trebled, but even at a rate of 1000 nucleotides per day the major proportion of time is spent managing the sequence data being generated. Hence, future improvements in gel reading and data entry procedures [35] or in the computer software that manipulates the data to assemble the final sequence [7-9] can be expected to contribute markedly to the overall speed with which DNA may be sequenced.

An obvious disadvantage of DNA sequencing by the shotgun method is the asymptotic nature of sequence accumulation (Fig. 3). This becomes an acute problem when the sequence is almost complete because the probability of picking a clone that comes from a region not already sequenced is greatly diminished. Much of the redundant sequence generated in the latter part of a project, however, serves the important function of confirming previous sequence, so it is useful to continue even when the overall rate of new sequence accumulation has dropped to a low level. Nevertheless, for some sequencing projects a more direct approach, in which specific regions are selected for sequence determination, may be the most appropriate way to fill in the final few gaps.

ACKNOWLEDGEMENTS

I would like to thank Dr. F. Sanger for his advice and encouragement during the course of this work and for a critical reading of the manuscript. Gifts of T4 DNA ligase from Dr. A.R. MacLeod, M13mWJ22(amber) RF DNA from Dr. G.P. Winter, and strain JM101 from Dr. J. Messing are also gratefully acknowledged. The author was supported by a Research Fellowship from the Cystic Fibrosis Foundation.

REFERENCES

1. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
2. Schreier, P.H. and Cortese, R. (1979) J. Mol. Biol. 129, 169-172.
3. Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) Nucleic Acids Res. 8, 1731-1743.
4. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
5. Heidecker, G., Messing, J. and Gronenborn, B. (1980) Gene 10, 69-73.
6. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
7. Staden, R. (1979) Nucleic Acids Res. 6, 2601-2610.
8. Gingeras, T.R., Milazzo, T.J.P., Sciaky, D. and Roberts, R.J. (1979) Nucleic Acids Res. 7, 529-545.
9. Staden, R. (1980) Nucleic Acids Res. 8, 3673-3694.
10. Messing, J., Gronenborn, B., Mueller-Hill, B. and Hofschneider, P.Y. (1977) Proc. Nat. Acad. Sci. USA 74, 3642-3646.
11. Gronenborn, B. and Messing, J. (1978) Nature 272, 375-377.
12. Cordell, B., Bell, G., Tischer, E., De Noto, F.M., Ullrich, A., Pictet, R., Rutter, W.J. and Goodman, H.M. (1979) Cell 18, 533-543.
13. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
14. Young, I.G. and Anderson, S. (1980) Gene 12, 257-265.
15. Anderson, S. and De Pamphilis, M.L. (1979) J. Biol. Chem. 254, 11495-11504.
16. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D.K. and Efstratiadis, A. (1978) Cell 15, 687-701.
17. Rothstein, R.J., Lau, L.F., Bahl, C.P., Narang, S.A. and Wu, R. (1980). In Methods in Enzymology, eds. Colowick, S.P. and Kaplan, N.C. (Academic Press, Inc., New York), Vol. 68, pp. 101-110.
18. Messing, J. (1979) Recombinant DNA Technical Bulletin 2, 43-48.
19. Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L. and Treiber, G. (1970) Virology 40, 734-744.
20. Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmas, R.C. (1981) Nucleic Acids Res. 9, 1691-1706.
21. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
22. Melgar, E. and Goldthwait, D.A. (1968) J. Biol. Chem. 243, 4409-4416.
23. Weiss, B., Jacquemin-Sablon, A., Live, T.R., Fareed, G.C. and Richardson, C.C. (1968) J. Biol. Chem. 243, 4543-4555.
24. Ehrlich, S.D., Bertazzoni, U. and Bernardi, G. (1973) Eur. J. Biochem. 40, 143-147.
25. Bernardi, A., Gaillard, C. and Bernardi, G. (1975) Eur. J. Biochem. 52, 451-457.
26. Sutcliffe, J.G. (1978) Nucleic Acids Res. 5, 2721-2728.
27. Lutter, L.C. (1977) J. Mol. Biol. 117, 53-69.
28. Heffron, F., So, M. and McCarthy, B.J. (1978) Proc. Nat. Acad. Sci. USA 75, 6012-6016.
29. Berkner, K.L. and Folk, W.R. (1977) J. Biol. Chem. 252, 3176-3184.
30. Westergaard, O., Brutlag, D. and Kornberg, A. (1973) J . Biol. Chem. 248, 1361-1364.
31. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457-465.
32. Shenk, T.E., Carbon, J. and Berg, P. (1976) J. Virol. 18, 664-671.
33. Frischauf, A.M., Garoff, H. and Lehrach, H. (1980) Nucleic Acids Res. 8, 5541-5549.

34. Challberg, M.D. and Englund, P.T. (1980) In Methods in Enzymology, eds. Colowick, S.P. and Kaplan, N.C. (Academic Press, Inc., New York) Vol. 65, pp. 39-43.
35. Gingeras, T.R. and Roberts, R.J. (1980) Science 209, 1322-1328.